Mozilla's Response to the National Telecommunications and Information Administration (NTIA) Request for Comments on AI Accountability Policy

June 2023

Table of Contents

About Mozilla

Mozilla's mission is to ensure the internet is a global public resource, open and accessible to all. An internet that truly puts people first, where individuals can shape their own experience and are empowered, safe and independent.

Founded as a community open source project in 1998, Mozilla currently consists of two organizations: the non-profit Mozilla Foundation, which leads our movement building work; and its wholly owned subsidiary, the Mozilla Corporation, which leads our market-based work, including the development of the Firefox web browser. The two organizations work in close concert with each other and a global community of tens of thousands of volunteers under the single banner: Mozilla.

For the past five years, Mozilla has been committed to advancing trustworthy AI. As we have noted in the 2020 white paper *Creating Trustworthy AI*, critical steps are necessary to make AI more trustworthy and ensure that it does not further deepen existing inequalities. This is why the Mozilla Foundation has been dedicating 100% of its $30M a year budget to philanthropic activities and advocacy on this topic. Further, Mozilla is investing another $30M in research and development on trustworthy AI via Mozilla.ai as well $35M in responsible tech startups — including startups with a focus on trustworthy AI — through Mozilla Ventures.

As an independent and mission-driven organization, Mozilla is committed to working with regulators to develop effective policies that balance the public interest with innovation and growth in the tech sector.

Introduction

Regulatory accountability mechanisms are commonplace in a wide variety of sectors, from the financial sector to food and drugs to the automotive sector. The AI industry, so far, is lacking such accountability mechanisms — despite the fact that it is rapidly expanding into an increasing number of areas in our lives, and into increasingly sensitive areas, too. At the same time, the harms that can arise when AI is used are well-established. Against this backdrop, the mantra of "move fast and break things'' back to which many AI companies currently seem to be reverting following the ascent of “generative AI” can no longer serve as the blueprint for product development and marketing. Imagine if drug or car manufacturers adopted the same motto.

A systematic approach to AI accountability is, therefore, overdue. Much important work is already underway when it comes to helping companies ensure that their technologies are developed and deployed responsibly. For instance, the White House recently announced several initiatives to promote responsible AI. Additionally, NIST’s AI Risk Management Framework as well as the White House OSTP’s Blueprint for an AI Bill of Rights, both of which Mozilla weighed in on, list a broad range of important considerations and concrete measures for actors along the AI value chain.

Still, these are only voluntary instruments. They help companies developing and deploying AI hold themselves accountable but do not mark significant progress towards companies being held accountable by others. What is needed now are rules that change the incentive structure in the AI industry and make companies price in the adverse effects on people, communities, and society caused by their products (Question 2).

AI assurance can play a significant role in this endeavor. The NTIA’s request for comments on this issue therefore comes at a timely moment and Mozilla welcomes the opportunity to share its views and experiences from our own activities in this field. In the following, we will highlight some important considerations regarding auditing as a mechanism for AI accountability and sketch out what a potential approach to fostering more accountability in this space could look like.

The promise and limitations of AI auditing

AI audits can be an important piece of the puzzle when it comes to fostering more accountability in the AI industry. Over the past years, adversarial audits conducted by independent researchers, investigative journalists, and civil society organizations have brought to light many of the harms that can be caused by AI. However, such audits are often bespoke, narrow projects specifically focused on one application or class of AI systems, led by small teams or individuals at organizations that lack the resources and access to replicate such investigations at scale. While this work can be effective in identifying or highlighting significant issues of concern, they generally lack the scope or scale to drive systemic or substantial changes.

While an AI auditing ecosystem is slowly emerging, it is still nascent and lacking standards for what constitutes a rigorous audit, both technically and procedurally. At Mozilla, we are committed to playing our part in fostering the development of a robust and trustworthy auditing ecosystem that can effectively uncover AI risks and subsequent harms, and point to pathways for prevention and mitigation. For example, we fund original research on the status quo in AI auditing as well as initiatives developing new tools and we run our own crowdsourced investigations of online platforms.

Public policy, too, can — for instance, by providing funding or creating demand for commissioned audits — stimulate the growth of this ecosystem and accelerate the development of standards and good practices, but it also must be careful not to over-rely on it as a lever for accountability. For instance, poorly implemented auditing mandates could create the risk of AI audits turning into performative check-box exercises that shield companies from criticism but accomplish little in the way of preventing harm and reining in risks.

Specifically, we want to note several aspects to take into consideration in designing public policy relying on AI audits.

Example 1:
Audits in the EU's Digital Services Act

The EU's Digital Services Act (DSA), which was passed last year, sets out new rules for a wide range of online services and platforms, including so-called Very Large Online Platforms and Search Engines (VLOPs and VLOSEs, respectively). In doing so, the DSA adopts a layered approach to assurance and oversight, which also extends to recommender systems (i.e., ranking algorithms) used by online platforms. In brief, these interconnected layers are the following:

- An obligation for VLOPs and VLOSEs to conduct risk assessments and develop mitigation measures regarding "systemic risks" linked to their service
- An obligation for VLOPs and VLOSEs to commission independent audits to assess their compliance with the obligations imposed by the DSA

- An obligation for VLOPs and VLOSEs to share data with vetted researchers to enable independent study of systemic risks linked to their service

First, the devil is in the details. Any regulatory auditing framework should consider diverse methodologies and be clear about the scope and focus of audits while accounting for varying contexts. For instance, who or what are the audit targets? This could range from entire companies to complex systems of AI models to individual models. It is also critical to keep in mind that an audit need not focus only on a technical artifact; it can — and in many cases should — also take into account the socio-technical and organizational context in which an AI system is developed and used (Question 1). To illustrate, even if adopting a more narrow scope, there's a vast difference between auditing one bespoke AI solution, for example a resume screening algorithm, a general-purpose large language model, such as OpenAI's GPT-4 or Google's Bard 2, or the complex suite of AI models powering the experience on social media or content-sharing platforms such as Instagram or YouTube. At the same time, audits can not only focus on specific models but also on the datasets used to train them (Question 5). All of these likely require different methods, tools, and resources. At the same time, auditing an entire service offered by a company like Google or Meta — similar to what is required under the EU's Digital Services Act (DSA) — may go far beyond the capacity of one individual auditing service and may require larger consortia to work together. For more complex audits, this would also raise the challenge of how to identify or prioritize different audit targets, e.g., one specific AI system integrated into a larger service (for example, an ad delivery system as opposed to an organic content-ranking system). Additionally, the more degrees of separation between the audit target and the dimension of risk (e.g., adverse impacts on democratic processes as opposed to directly measurable consequences for an individual interacting with an AI system) the more difficult to establish causal links between AI systems and higher-level impacts of an AI system (Question 4). Still, a breadth of assessments and different methodologies — including more qualitative methods — may be needed to also assess these more systemic or collective risks.

Second, it is important to untangle incentives in the auditing ecosystem — only where the incentive structure is right and auditors are sufficiently independent (and have sufficient access) can there be more certainty that audits aren't simply conducted for the purpose of "audit-washing" (Question 7). Therefore, any auditing framework must ensure that the incentives it creates for auditors and other stakeholders are not misaligned with regulators' intention. For instance, to ensure independence, it is critical to take into account potential conflicts of interests on the side of auditors. Where auditors have a commercial interest in producing a positive audit report, the integrity of the audit is fundamentally compromised. In its independent auditing obligation for very large online platforms and search engines (Article 37), the DSA, for instance, outlines several requirements for auditors in order to ensure auditors' independence (Question 14):

- The auditor may not provide non-auditing services to the audited company for 12 months prior to or following the audit
- The auditor may not provide auditing services to the audited company for more than 10 years
- Fees paid to auditors cannot be contingent on the result of the audit

Still, even such requirements do not fully eliminate potential conflicts of interest: For example, being able to be commissioned to conduct audits for 10 consecutive years still offers strong financial incentives to auditors not to "sour" a business relationship with a platform. Similarly, the DSA prescribes nothing about a potential "revolving door" between platforms and auditors, where auditors may be hired by platforms after collaborating.

The DSA's tiered approach to AI assurance further raises another important consideration: the question of who performs audits. This is not limited to professional auditing services commissioned by companies themselves (or, potentially, by other actors such as regulators). To this point, critical advancements in auditing, and many of the most-cited examples of audits, have been made by researchers, journalists, and civil society organizations. However, it cannot be the solution to outsource accountability to actors who may already be under-resourced or who may need to engage in such activities because the groups they represent themselves are most at risk of harm. Therefore, relying on adversarial audits from such groups should be one piece of the puzzle, yet cannot serve as the foundation of accountability in the AI ecosystem. At the same time, creating an enabling environment for such adversarial audits is crucial — in many cases, having to rely on companies' cooperation can constrain auditors or function as a bottleneck in performing audits. For instance, our own crowdsourced research on YouTube has helped us uncover significant flaws in YouTube's ineffective user controls. Similarly, many other researchers also rely on crowdsourced or scraped data or perform so-called "sock-puppet audits" (i.e., creating accounts to simulate user journeys) to carry out adversarial audits. Regulation should enable such public-interest researchers — with robust security and privacy protocols in place — instead of allowing such research to be chilled due to fear of legal liability (e.g., due to terms of service violations).

Mozilla supports the Platform Transparency and Accountability Act (PATA), sponsored by Sen. Coons (D-DE) and cosponsored by Sens. Cassidy (R-LA), Klobuchar (D-MN), Cornyn (R-TX), Blumenthal (D-CT), and Romney (R-UT), which is a good example of policy that would enable true platform transparency via researcher access. PATA is a bipartisan proposal that would require social media platforms to provide access to data for public-interest research projects, and create valuable ad transparency. It would also establish a legal "safe harbor" that would enable legitimate public-interest research, free from threats of legal action (as we do in our Bug Bounty programs), while protecting privacy and security.

Example 2:
Conformity assessments in the EU's draft Artificial Intelligence Act

The EU's draft AI Act — which is currently being negotiated by the EU institutions — proposes horizontal rules for AI that would apply across sectors. Most importantly, these rules would apply to high-risk applications made available on the European market (and potentially so-called "general-purpose AI" or "foundation models"). Developers of such high-risk applications would need to (self-)assess conformity with a set of requirements prescribed by the AI Act, including requirements around risk management, accuracy and robustness, data governance, and human oversight. Critically, these high-level requirements of the AI Act are likely to be operationalized through standards developed by European standard-setting organizations and formally recognized by the European Commission following adoption of the AI Act.

Further, it is important to remember who develops the tools, benchmarks, and standards that play an important role in audits and assessing aspects like accuracy or robustness. Regulators should thus be wary of uncritically adopting industry benchmarks or standards — even if they are in widespread use — and should ensure that independent experts are heard in the process of developing rules or guidance in this context. Benchmarks, for example, are often developed by companies themselves and may be narrow in what exactly they assess. The state of the art is fluid here and certain benchmarks may become rapidly outdated. Further, they may only be of use to assess a small subset of relevant aspects.

With regard to the role of standards, the case study of the EU's proposed AI Act — which is currently being negotiated — is an illustrative example (Question 14). To assess (or for companies to self-assess) whether a high-risk AI system made available in the EU is compliant with the requirements prescribed by the AI Act, standards developed by European standard-setting organizations are likely to become the key measuring stick. However, as has been noted by academic researchers and civil society, these standards could also extend to aspects concerning people's fundamental rights — as opposed to solely technical aspects, as is usually the norm for standards. Further, representatives from civil society organizations and groups representing particularly vulnerable or marginalized communities have historically been underrepresented in the standard development process and, where they are represented, often lack the resources for deeper engagement. This highlights the importance of relying on standards and benchmarks in audits of AI systems that also consider the interests of those most likely to be affected and harmed by AI systems — and to take into account in developing auditing mandates what standards and benchmarks are already available or would need to be developed in the future.

Finally, taking into account both the AI value chain and lifecycle is crucial. Regulators need to consider at what point in the AI value chain and lifecycle harms originate, and where they can best be addressed. With regard to the value chain, different types of harms may be rooted at different points in the value chain and would need to be mitigated by different actors (Question 15). For example, harmful biases in the data used to pre-train an AI model made available for integration via an API can much easier be identified and addressed upstream rather than downstream by those integrating the model into a product or service. At the same time, some risks arising from the use of AI are highly context-specific and may be best assessed by downstream actors at the application layer.

With regard to the lifecycle, audits may be required at different points in time as aspects such as performance and risks might change over time, so that one-off audits at any one point may not suffice (Question 16). In many cases, audits may thus be necessary with a regular cadence (the DSA, for instance, requires large-scale independent audits of very large online platforms at least once a year).

In summary, regulators need to consider a variety of challenges in developing a robust framework for AI accountability, and auditing specifically: ensuring auditors' independence and creating an incentive structure that is aligned with enhancing trustworthiness and accountability in AI; enabling adversarial audits and public interest researchers while not delegating the work of advancing AI accountability to them entirely; critically examining the role, usage, and origin of benchmarks, standards, and other tools; and considering the origins of harm and suitable points of intervention along the AI value chain and throughout the lifecycle of AI systems.

Toward a comprehensive approach to AI accountability

With these considerations on AI assurance — and auditing in particular — in mind, what could a comprehensive approach to AI accountability look like?

In the absence of comprehensive horizontal legislation on AI at the federal level, a vertical (or sectoral) approach can still effectively serve the goal of ensuring accountability across the AI ecosystem. In this context, regulatory audits can serve both as a complement to internal, commissioned, or adversarial audits or advance accountability independently of such practices. In summary, this would comprise four key aspects: Strengthening regulatory capabilities and authorities; creating a "paper trail" on the side of entities developing or deploying AI; creating a legal basis for enhanced accountability where potential AI-enabled harms are not yet covered by existing legislation; and aiding regulators in the discovery of potential harms and targets for regulatory audits.

First, it is important to acknowledge that there already are areas in which regulators and government oversight bodies have the mandate to hold entities accountable for potential harms caused by AI or misuse of AI. For instance, in a recent joint letter, the Consumer Financial Protection Bureau (CFPB), Department of Justice (DOJ), Equal Employment Opportunity Commission (EEOC), and Federal Trade Commission (FTC) provide an overview of their enforcement powers and activities in this area, for example regarding discriminatory lending, housing, or employment decisions. However, such a mandate is a necessary but not sufficient condition to ensure regulatory oversight and, ultimately, accountability.

To enable enforcement bodies to fulfill their mandates, they also need adequate resources and capabilities, including the necessary technical — and interdisciplinary — expertise and tooling to perform regulatory audits or inspections. Regulators in the United States, such as the FTC, struggle to keep abreast of the current AI ecosystem with the meager funding they are given. Increasing the funding of these agencies would allow regulators to dedicate more energy and resources to ensuring AI accountability across domains. Just as Congress has acknowledged[1] that it costs more to recruit financial-sector experts to government service (by implementing a different payscale at financial regulatory agencies), it should acknowledge that an accountable AI ecosystem requires agency funding commensurate with the task of watching over the technology industry (Question 12).

Further, regulators' authorities may need to be strengthened in order for them to be able to perform more robust and diligent audits. For example, where not already existing, this may include extending (more far-reaching) administrative subpoena powers to regulators, for example in order to subpoena technical documentation, documents, or technical access to systems for testing and evaluation purposes. The efficacy of such mechanisms can further be strengthened by creating a "paper trail",

[1] As one GAO report describes: "Congress granted the federal financial regulatory agencies the flexibility to establish their own compensation systems recognizing that the existing approach to compensating employees could impede these agencies' ability to recruit and retain employees critical to meeting their organizational missions."

and “auditability” more broadly, that regulators can rely on. For instance, sectoral regulators could compel developers and/or deployers of AI systems — or of certain AI systems, based on predefined criteria — to keep basic technical documentation about AI design, development, and deployment, including model and dataset documentation (also with regard to data provenance, collection, and curation) such as model cards or datasheets (Question 20). Such documentation could potentially help regulators as a starting point in their investigations.

Other lower-level and sector- or agency-specific policy changes can similarly have outsized effects on the ecosystem (Question 30). For example, in 2020 the FTC lost its ability to obtain consumer refunds in court for many unfair and deceptive practices, and its ability to obtain injunctions for past conduct is severely threatened. If, say, the FTC finds that a company has made deceptive claims in its algorithmic audits, the FTC’s options for consumer redress are now significantly limited. Congress has the power to restore this crucial power. The FTC’s example demonstrates that seemingly unrelated agency-specific policy choices can have a follow-on effect when it comes to AI accountability.

Furthermore, while many agencies already have a mandate that also applies to AI, not all AI-enabled harms are covered by existing legislation or other rules (Question 26). For example, potential harms to children exposed to AI systems in the education context or risks to workers stemming from algorithmic management — for example of warehouse workers or in the gig economy — warrant further protections and accountability measures. Regulators in these and other areas with similar gaps concerning AI should be given a statutory basis for enforcement to protect people from harm. This could include empowering regulators to designate particularly risky categories of AI systems or contexts of use and/or to adopt domain-specific obligations for developers and deployers, for instance testing, documentation, or rigorous external audits performed by independent auditors in the case of AI systems used in sensitive contexts (Question 18).

To further strengthen accountability, mechanisms that aid regulators in identifying potential harms or targets for investigation should also be considered. This could include the introduction of formal complaint mechanisms for individuals and organizations representing their interests to lodge complaints with regulators if there is cause to believe that they may have been harmed by an AI system. Especially if tied to effective transparency measures, these mechanisms could bring cases of potentially harmful uses of AI to regulators' attention and point them into promising directions for prospective investigations.

Finally, the lack of a federal comprehensive federal privacy law is another major barrier to AI accountability (Question 25). It is important to recognize that not all AI systems will be trained on, or otherwise use, personal information or even de-identified personal information. Furthermore, an AI chatbot might be trained on publicly available information (which would not be covered by many privacy legislative proposals), and such a system might not collect or store user queries at all. Additionally, an AI model trained on personal information may or may not be considered personal information on its own, depending on its propensity to leak personal information, among other factors.

Nevertheless, as the AI ecosystem grows, it is likely that an increasing number of AI systems will rely on the collection, processing, or transfer of data in ways that would be regulated by a federal comprehensive privacy law. For example, an AI-based product recommendation system on an e-commerce website is typically premised on the collection of consumer shopping behavior. A federal comprehensive data privacy law would give consumers control over the data that AI systems can collect, process, and transfer.

At the same time, comprehensive privacy rules can also be an important piece in strengthening AI assurance and auditing by ensuring that auditors — both internal and external — must respect the privacy and security of users of a service or people affected by an AI system in assessing said system, including in the case of regulatory audits by government bodies. Protecting, for example, personal information in the course of audits can both enhance the legitimacy of AI assurance practices and increase legal certainty among auditors when it comes to necessary security protocols.

For these (and other) reasons, Mozilla supports the passage of the American Data Privacy and Protection Act ("ADPPA"), which, in addition to providing essential data rights, has specific provisions related to AI. For example, the bill prohibits AI-based discrimination that uses "covered data," and would require certain impact assessments and design evaluations. The ADPPA does carve out certain sectors (notably portions of the financial and health sectors) that frequently engage in AI-based decision-making; for these sectors, continued regulatory and legislative focus will be essential.

Conclusion

Mozilla supports the NTIA's effort to advance the national conversation about AI accountability policy. Raising the bar in this respect is a necessary step. AI assurance and auditing hold much promise as a key component of such efforts. However, regulators should also be mindful of the limitations of AI assurance and auditing in order to ensure that such mechanisms don't turn into inconsequential and performative check-box exercises. Instead, they should pay close attention to aspects such as independence, auditability, and benchmarks or standards. Further, the NTIA should also consider the role of regulatory audits in a comprehensive approach to AI accountability, strengthening regulators' capacity to perform audits and empowering regulators who so far lack the mandate to protect people from harm.

If designed well, AI accountability policy can ensure that AI serves the interest of all people and move us towards more trustworthy AI while at the same time enabling more purpose-driven innovation and growth in the tech sector. As an independent and trusted voice in the tech sector, Mozilla stands ready to support these efforts.

Mozilla appreciates the opportunity to comment and to provide our perspective as both a non-profit foundation and a technology company. We would be happy to further discuss any questions you may have regarding our comments.